Available online at www.sciencedirect.com
ScienceDirect
Discrete Mathematics 308 (2008) 5771–5781
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

Estimates of the Pythagoras number of $\mathbb{R}_m[x_1, \ldots, x_n]$ through lattice points and polytopes

David B. Leep, Colin L. Starr*

Department of Mathematics, University of Kentucky, Lexington, KY 40506-0027, USA
Department of Mathematics, Willamette University, Salem, OR 97301, USA

Received 18 July 2005; received in revised form 22 August 2007; accepted 30 October 2007
Available online 20 December 2007

Abstract

Hilbert's 17th Problem launched a number of inquiries into sum-of-squares representations of polynomials over the real numbers. Choi, Lam, and Reznick gave some bounds on the number of squares required for such a representation and indicated some directions for improving these bounds. In the first part of this paper, we follow their suggestion and obtain some stronger bounds. In the second part, we show that in the case of homogeneous polynomials in three variables, this technique cannot be extended further.

Keywords: Pythagoras number; Cage; Sum of squares; Lattice points; Polytopes

1. Introduction

Let R be a commutative ring with multiplicative identity. Assume that $a \in R$ is a sum of squares in R: $a = \sum a_i^2$, where each $a_i \in R$. We wish to know how many squares are required for such a representation. Accordingly, we have the following definition.

Definition 1.1. If a is a sum of squares in R, then the *length* of a in R is defined by

$$\operatorname{length}(a) = \min \left\{ t \;\middle|\; a = \sum_{i=1}^{t} a_i^2 \text{ with } a_i \in R \right\}.$$

Example 1.2. In $\mathbb{Z}$, length$(6) = 3$.

Furthermore, if we are given a subset S of R, we want to know the maximum length of an element of S, if there is such a bound.

* Corresponding author at: Department of Mathematics, Willamette University, Salem, OR 97301, USA.
E-mail addresses: leep@ms.uky.edu (D.B. Leep), cstarr@willamette.edu (C.L. Starr).

0012-365X/$ - see front matter
doi:10.1016/j.disc.2007.10.045

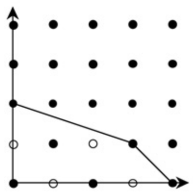

Fig. 1. $C(f)$.

Definition 1.3. Let $S \subseteq R$. The *Pythagoras number* of S is

$$P(S) = \sup\{\text{length}(a) | a \text{ is a sum of squares in } R\}.$$

Example 1.4 (*Lagrange*). $P(\mathbb{Z}) = 4$.

Example 1.5 (*[3], [4, p. 302]*). $P(\mathbb{R}[x]) = 2$.

Example 1.6. $P(\mathbb{R}[x_1, \ldots, x_n]) = \infty$ for $n \geq 2$. [1, Theorem 4.1, p. 56]

The ring $\mathbb{R}[x_1, \ldots, x_n]$ is of special interest. Instead of considering all of $\mathbb{R}[x_1, \ldots, x_n]$, we will consider subsets of $\mathbb{R}[x_1, \ldots, x_n]$ obtained by restricting the degrees of the polynomials.

Definition 1.7. $\mathbb{R}_m[x_1, \ldots, x_n]$ denotes the set of all homogeneous polynomials of degree m in n variables; we refer to these as n-ary m-ics. We are primarily interested in even m.

Choi, Lam, and Reznick published some bounds on $P(\mathbb{R}_m[x_1, \ldots, x_n])$ [2]. We will exploit their methods to improve the estimates for this Pythagoras number in Section 2 and discuss in Section 3 some of the problems that limit the use of their methods. We first introduce the idea of *cages* as presented in [2,7] and expanded upon in [8].

Definition 1.8 (*[2,7]*). Let $f \in \mathbb{R}_m[x_1, \ldots, x_n]$; $f = \sum_{|\alpha|=m} a_\alpha x^\alpha$ in multi-index notation, where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n$, each $\alpha_i \geq 0$, and $|\alpha| = \alpha_1 + \cdots + \alpha_n$. The *cage* of $f, C(f)$, is the closed convex hull of $\{\alpha | a_\alpha \neq 0\} \subseteq \mathbb{R}^n$.

The idea is to translate the algebraic problem into a geometric one; the example below illustrates this. Note that in Example 1.9, the polynomial is not homogeneous. If we were to homogenize f, the cage of f would lie in $\mathbb{R}^3$; however, that cage would be a plane in $\mathbb{R}^3$, and thus could be projected into $\mathbb{R}^2$ without losing any information. Rather than homogenizing and projecting the cage, we just plot those points (a, b) in $\mathbb{R}^2$ corresponding to monomials $x^a y^b$ occurring in f. The two approaches are equivalent.

Example 1.9. The cage of $f(x, y) = 1 + x^2 + 3xy + y^2 - x^4 + x^3y$ appears in Fig. 1.

The cage itself is the region enclosed by the quadrilateral together with the quadrilateral. The solid points plotted on and in the quadrilateral are (0, 0), (2, 0), (1, 1), (0, 2), (4, 0), and (3, 1), corresponding to the monomials 1, x^2, $3xy$, y^2, x^4, and x^3y, respectively.

Definition 1.10. $C_{n,m} = \{(c_1, \ldots, c_n) | 0 \leq c_i \in \mathbb{R} \text{ and } \sum c_i = m\}$ is referred to as the *full cage*.

The simplex $C_{n,m}$ represents the cage of a generic n-ary m-ic; if $f \in \mathbb{R}_m[x_1, \ldots, x_n]$, then $C(f) \subseteq C_{n,m}$. It is a subset of a hyperplane in $\mathbb{R}^n$.

One of the many useful aspects of the geometric approach to sums-of-squares problems is the simple relationship between the cage of a polynomial and the cage of its square.

Example 1.11. In Fig. 2, the cage of

$$f^2 = 1 + 2x^2 + 6xy + 2y^2 - x^4 + 8x^3y + 11x^2y^2 + 6xy^3 + y^4 - 2x^6 - 4x^5y + 4x^4y^2 + 2x^3y^3 + x^8 - 2x^7y + x^6y^2$$

is superimposed on the cage of $f = 1 + x^2 + 3xy + y^2 - x^4 + x^3y$.

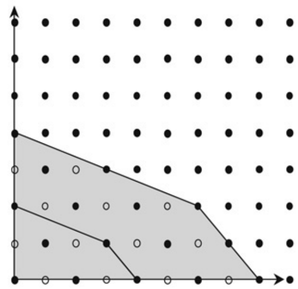

Fig. 2. $C(f^2)$.

The cage of f^2 is the "doubled cage" of f; it is similar to $C(f)$ with a scale factor of two and has sides parallel to those of $C(f)$ [7, Lemma to Theorem 1].

The method used in [2] requires that we know how many lattice points (of certain kinds) appear in $C(f)$; we adopt their notation.

Definition 1.12. Let C be a cage; that is, let $C = C(f)$ for some polynomial f. Then $L(C) = C \cap \mathbb{Z}^n$ (lattice points in C), $E(C) = C \cap (2\mathbb{Z})^n$ (even lattice points in C), and

$$A(C) = \left\{ \alpha \,\middle|\, \alpha = \frac{\beta + \beta'}{2} \text{ for some } \beta, \beta' \in E(C) \right\}$$

(averages of even lattice points in C). Also, $a(C) = |A(C)|$ and $e(C) = |E(C)|$.

For the full cage $C_{n,m}$, $a(C_{n,m}) = \binom{n+m-1}{n-1}$ [2, Lemma 3.4] and $e(C_{n,m}) = \binom{n+m/2-1}{n-1}$. For any cage C, $a(C) \le \binom{e(C)+1}{2}$.

Now consider the following theorem of Reznick:

Theorem 1.13 *([7, Lemma to Theorem 1]). If $f \in \mathbb{R}[x_1, \ldots, x_n]$ is positive semi-definite (psd), then $C(f)$ is the convex hull of $E(C(f))$. Conversely, if $C \subseteq C_{n,m}$ is the convex hull of a set of even lattice points, then there exists a psd polynomial f such that $C(f) = C$.*

In the light of this theorem, we may consider $P(C)$, the Pythagoras number of the set of polynomials having a cage contained in C. In particular, $P(C_{n,m}) = P(\mathbb{R}_m[x_1, \ldots, x_n])$. This brings us to the main result of [2].

Theorem 1.14 *([2, Theorems 4.4 and 6.1]). Let C be a cage, let $e = e(C)$, and let $a = a(C)$. Then*

$$\frac{a}{e} \le \lambda(C) = \frac{2e + 1 - \sqrt{(2e+1)^2 - 8a}}{2} \le P(C) \le \Lambda(C) = \frac{\sqrt{1 + 8a} - 1}{2} \le e.$$

Thus, $\lambda(C)$ is a lower bound and $\Lambda(C)$ is an upper bound for $P(C)$, and both are obtained from counting the number of even lattice points and the number of distinct averages in a cage. Because we have precise formulas for $e(C_{n,m})$ and $a(C_{n,m})$, the above theorem gives specific upper and lower bounds for $P(C_{n,m})$. In fact, if $m = 2$, then $e = n$ and $a = \frac{n(n+1)}{2}$, so $\lambda(C) = n$ and $\Lambda(C) = n$. Thus, in this case, $P(C) = n$.

Notice that if $C \subseteq C_{n,m}$, then $P(C) \le P(C_{n,m})$. However, it is not necessarily the case that $\lambda(C) \le \lambda(C_{n,m})$. It is possible that

$$P(C_{n,m}) \ge P(C) \ge \lambda(C) > \lambda(C_{n,m}).$$

This observation is the focus of this paper. We apply these results of [2] to improve the lower bound estimates for $P(C_{n,m})$ and consider the extent to which these methods may be exploited.

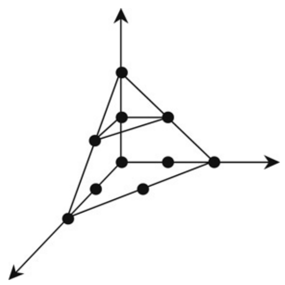

Fig. 3. The full cage $C(4, 4)$. $e(C_{4,4}) = 10$, $a(C_{4,4}) = 35$, $\lambda(C_{4,4}) = 4.1555$. •: Even point.

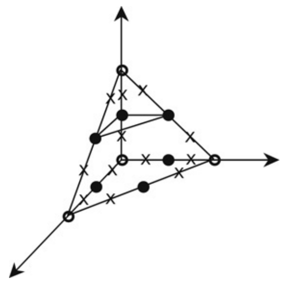

Fig. 4. "Clipped corners". $e(C') = 6$, $a(C') = 19$, $\lambda(C') = 4.4384$. ∘: Deleted even point. •: Even point. x: Lost average.

2. Improving lower bound estimates of $P(C_{n,m})$

In [2], a family of subcages $C \subseteq C_{n,m}$ is given that satisfies $\lambda(C) > \lambda(C_{n,m})$, as follows:

Example 2.1 (*[2, p. 124]*). First,

$$\lambda(C_{n,4}) = \left(\frac{1}{2} - \frac{1}{\sqrt{6}}\right) n^2 + \frac{1}{2} n + \frac{1}{2}\left(1 + \frac{1}{\sqrt{6}}\right) + o(1).$$

See Fig. 3 for a rendering of the full cage $C_{4,4}$.

Let C' be the subcage of $C_{n,4}$ obtained by taking the convex hull of the set of even lattice points

$$E(C') = \{2e_i + 2e_j | 1 \le i < j \le n\},$$

where e_i refers to the ith standard basis vector. Then [2, p. 125] shows that

$$\lambda(C') = \left(\frac{1}{2} - \frac{1}{\sqrt{6}}\right) n^2 + \left(\frac{3}{\sqrt{6}} - \frac{1}{2}\right) n + \left(\frac{1}{2} - \frac{1}{\sqrt{6}}\right) + o(1).$$

The cage for C' is shown in Fig. 4. In [2] this is referred to as "clipping the corners" of $C_{n,m}$: all vertices of the form $4e_i$ are missing, and the convex hull is taken over the remaining even lattice points.

This yields a dramatic improvement over the full cage estimate:

$$\lambda(C') - \lambda(C_{n,4}) = \left(\frac{3}{\sqrt{6}} - 1\right) n - \frac{3}{2\sqrt{6}} + o(1) \to \infty$$

as $n \to \infty$.

This example motivates the following definition.

Definition 2.2 (*[2, p. 125]*). $\overline{\lambda}(C_{n,m}) = \sup\{\lambda(C) | C \subseteq C_{n,m}\}$.

Determining $\overline{\lambda}$ will indicate how far this method of estimating $P(C_{n,m})$ can be exploited.

2.1. Strategy

We construct subcages of $C_{n,m}$ that satisfy $\lambda(C) > \lambda(C_{n,m})$ and that also satisfy $\lambda(C) > \lambda(C')$, where C' is the clipped-corners cage of Example 2.1.

Lemma 2.3. *Let C be an n-dimensional cage with $e(C)$ even points and $a(C)$ distinct averages of even points. Let $P \in C$ be a vertex (corner) of C, and let $\hat{C}$ be C with P deleted. Then the closed convex hull of $\mathbb{Z}^n \cap \hat{C}$ contains at most $a(C) - n - 1$ distinct averages of even points.*

Proof. Since C is an n-dimensional polytope, it follows from [9, Chapter 2], that P belongs to at least n edges.

Along each edge is some nearest even lattice point; the midpoint of the segment joining P to this point is an average of two even points. Because P is a vertex of the cage, deleting P will also delete all such averages—they only appear as averages of P with some other point. (That is, those averages are not duplicated by other pairs of even lattice points.) Finally, since P itself (the average of P and P) is lost, the total number of deleted averages must be at least $n + 1$. □

A full-dimensional cage in $\mathbb{R}^n$ is an $(n-1)$-dimensional object, so deleting a vertex will also delete *at least* $(n-1)+1 = n$ distinct averages.

In order to create a subcage of a given cage, we delete one or more vertices from the given cage. We can also delete several in succession, but the deletions must be performed in the correct order: an interior point P of an edge may become a vertex after some other vertex is deleted, and then P may be deleted as well. However, if P is deleted first, then P reappears upon taking the convex hull of what remains, and nothing has changed. This motivates the following definition.

Definition 2.4. An *allowable deletion* of a vertex from a cage C (of a psd polynomial) is the deletion of an even point P of a cage such that

(a) P is a vertex of the cage and
(b) the closed convex hull of the even points of $C - \{P\}$ has the same dimension as C.

The resulting cage is the closed convex hull of the even points of $C - \{P\}$. A sequence of deletions is performed in *correct order* if it is a sequence of allowable deletions.

In terms of improving the lower estimate, there is no loss in assuming that deletion of P does not reduce the dimension of the cage.

Theorem 2.5. *If deletion of a point P from a cage C reduces the dimension of the cage, then $\lambda(C - \{P\}) < \lambda(C)$.*

Proof. If $e = E(C)$ and $a = A(C)$, then $E(C - \{P\}) = e - 1$ and $A(C - \{P\}) = a - (e-1) - 1 = a - e$. Thus

$$\begin{aligned}\lambda(C - \{P\}) &= \frac{2(e-1)+1-\sqrt{(2(e-1)+1)^2 - 8(a-e)}}{2}\\ &= \frac{2e-1-\sqrt{(2e+1)^2-8a}}{2}\\ &< \lambda(C). \quad \square\end{aligned}$$

For convenience, let

$$a_{n,m} = \binom{n+m-1}{n-1} \quad \text{and} \quad e_{n,m} = \binom{n+m/2-1}{n-1}.$$

Lemma 2.6. *For any $n \geq 2, n \in \mathbb{Z}$ and $m \geq 2, m \in 2\mathbb{Z}$, we have $a_{n,m} - ne_{n,m} + \binom{n}{2} \geq 0$. Equality occurs if and only if $n = 2$ or $m = 2$ or $(m, n) = (4, 3)$.*

Proof. Let $g(n,m) = a_{n,m} - ne_{n,m} + \binom{n}{2}$. For $n = 2$, we have

$$g(2,m) = \binom{m+1}{1} - 2\binom{m/2+1}{1} + \binom{2}{2} = 0,$$

so the lemma holds (with equality) for any m when $n = 2$. We proceed by induction on n. The induction hypothesis then states that

$$g(n,m) = a_{n,m} - ne_{n,m} + \binom{n}{2} \geq 0$$

for some $n \geq 2$ and all even $m \in \mathbb{Z}^+$.

We wish to show that

$$\begin{aligned} g(n+1,m) &= \binom{(n+1)+m-1}{(n+1)-1} - (n+1)\binom{(n+1)+m/2-1}{(n+1)-1} + \binom{(n+1)}{2} \\ &\geq 0. \end{aligned}$$

Using the identity

$$\binom{r+1}{s+1} = \frac{r+1}{s+1}\binom{r}{s},$$

this becomes

$$\binom{n+m-1}{n-1}\frac{n+m}{n} - (n+1)\binom{n+m/2-1}{n-1}\frac{n+m/2}{n} + \binom{n+1}{2} \geq 0,$$

or simply

$$\frac{n+m}{n}a_{n,m} - (n+1)\frac{n+m/2}{n}e_{n,m} + \binom{n}{2} + n \geq 0.$$

The left-hand side may be rewritten as

$$\begin{aligned} &\frac{n+m}{n}a_{n,m} - n\frac{n+m}{n}e_{n,m} + n\cdot\frac{m}{2n}e_{n,m} - \frac{n+m/2}{n}e_{n,m} + \frac{n+m}{n}\binom{n}{2} - \frac{m}{n}\binom{n}{2} + n \\ &= \frac{n+m}{n}\left[a_{n,m} - ne_{n,m} + \binom{n}{2}\right] + e_{n,m}\left(\frac{nm-2n-m}{2n}\right) - m\left(\frac{n-1}{2}\right) + n \\ &= \frac{n+m}{n}\left[a_{n,m} - ne_{n,m} + \binom{n}{2}\right] + \frac{e_{n,m}}{n}\left(\frac{m}{2}(n-1) - n\right) - \left(\frac{m}{2}(n-1) - n\right) \\ &= \frac{n+m}{n}g(n,m) + \left(\frac{e_{n,m}}{n} - 1\right)\left(\left(\frac{m}{2} - 1\right)(n-1) - 1\right). \end{aligned} \tag{1}$$

We need to show that this is nonnegative. The first term on the left-hand side is nonnegative by the induction hypothesis, so it suffices to show that the second term is nonnegative for even $m \geq 2$. If $m = 2$, then

$$e_{n,m} = \binom{n}{n-1} = n,$$

so the second term is zero. We observe that $e_{n,m}$ is an increasing function in m because

$$e_{n,m+2} = \frac{n+m/2}{1+m/2}e_{n,m} > e_{n,m}.$$

Thus the first factor of the second term is nonnegative for even $m \geq 4$. It is also clear that the second factor of the second term is nonnegative for $m \geq 4$; thus $a_{n,m} - ne_{n,m} + \binom{n}{2} \geq 0$.

If $n = 2$ or $m = 2$ or $(n, m) = (3, 4)$, it is easy to compute that $a_{n,m} - ne_{n,m} + \binom{n}{2} = 0$. On the other hand, suppose that $n > 2$, $m \geq 4$, and $g(n, m) = a_{n,m} - ne_{n,m} + \binom{n}{2} = 0$. Since

$$g(n, m) = \frac{m + (n - 1)}{n - 1} g(n - 1, m) + \left(\frac{e_{n-1,m}}{n - 1} - 1\right)\left[\left(\frac{m}{2} - 1\right)(n - 2) - 1\right] = 0$$

from (1), and both terms in the middle expression are nonnegative, we see that both terms equal to zero. Thus either $e_{n-1,m} = n - 1$ or $\left(\frac{m}{2} - 1\right)(n - 2) = 1$. The former requires $m = 2$, which is not the case, so the latter must hold. Because this is an integer equation, we must have both $\frac{m}{2} = 2$ and $n - 2 = 1$, so $(n, m) = (3, 4)$. $\square$

Lemma 2.7. *Let $e = e(E(C_{n,m}))$ and $a = a(A(C_{n,m}))$. Let C_k be a cage obtained by allowably deleting k points from E, and put $e' = e(C_k) = e - k$, where $k \leq e - n$. Suppose that $a' = a(C_k) = a - nk$. If C_{k+1} is a cage resulting from the allowable deletion of one even point from C_k and $a(C_{k+1}) = a' - n$, then $\lambda(C_k) \leq \lambda(C_{k+1})$. If in addition, $m > 2$, $n > 2$, and $(m, n) \neq (4, 3)$, then $\lambda(C_k) < \lambda(C_{k+1})$.*

Proof. Let

$$f(k) = \frac{2(e - k) + 1 - \sqrt{(2(e - k) + 1)^2 - 8(a - nk)}}{2};$$

treating k as a continuous variable, we will show that $f(k) = \lambda(C_k)$ is increasing in k.

Because

$$f'(k) = -1 + \frac{2e - 2k + 1 - 2n}{\sqrt{(2(e - k) + 1)^2 - 8(a - nk)}},$$

it suffices to show that

$$2e - 2k + 1 - 2n \geq \sqrt{(2(e - k) + 1)^2 - 8(a - nk)}.$$

Since both sides are positive, this is equivalent to showing that

$$(2e - 2k + 1 - 2n)^2 \geq (2(e - k) + 1)^2 - 8(a - nk),$$

which reduces to just

$$a - ne + \binom{n}{2} \geq 0.$$

Thus, f is nondecreasing in k by the preceding lemma, and if $n > 2$, $m > 2$, and $(m, n) \neq (4, 3)$, then f is strictly increasing in k. $\square$

This means that the more points we can delete, that cost us no more than n distinct averages (as Lemma 2.3 says they must at minimum cost us), the larger the estimate λ becomes, and the closer to $P(C_{n,m})$. Also, it is easy to see that if the deletion of some even point results in the deletion of more than n averages, then the resulting λ will be smaller than if only n averages were deleted: the formula for λ shows that λ is increasing in $a(C)$. Thus, for a given cage C with $e - k$ even points, $\lambda(C)$ is no larger than the estimate obtained by deleting nk averages with k even points.

It is also important to observe that any $(n - 1)$-dimensional cage can be obtained from $C_{n,m}$ by a sequence of allowable deletions.

Theorem 2.8. *If C is an $(n - 1)$-dimensional cage of a psd polynomial, then C can be obtained from $C_{n,m}$ by a sequence of allowable deletions.*

Proof. Let C be a cage of an arbitrary psd n-ary m-ic. Then C is a subcage of $C_{n,m}$. The proof is by induction on the difference between $E(C)$ and $E(C_{n,m})$. If $C = C_{n,m}$, then the proof is complete, so assume that $C \neq C_{n,m}$ and that if C' is another subcage of $C_{n,m}$ with $E(C_{n,m}) - E(C') < E(C_{n,m}) - E(C)$, then C' is obtainable by a sequence of allowable deletions.

Since $C \neq C_{n,m}$, there is an even point in $C_{n,m} - C$. Let P be a nearest such even point; that is, let P be an even point of $C_{n,m}$ such that no other even points lie between P and C. Let C' be the closed convex hull of $C \cup \{P\}$. Then

C' is obtainable from $C_{n,m}$ by a sequence of allowable deletions. Since P is a vertex of C', and C and C' have the same dimension, deletion of P from C' is also allowable. Thus C is obtainable from $C_{n,m}$ by a sequence of allowable deletions. □

In the next section, we exploit Lemma 2.7.

3. Examples

Let r be the least nonnegative residue of n modulo 3, and let e_i be the ith standard basis vector.

Theorem 3.1. *Let C be the subcage of $C_{n,4}$ obtained by deleting all points of the forms*

$$4e_{3k}, \quad 4e_{3k+2}, \quad 2e_{3k+1}+2e_{3k+2}, \quad \text{and} \quad 2e_{3k+1}+2e_{3k+3},$$

with the following modifications:

$$\text{Delete} \quad \begin{array}{ll} 4e_n & \text{if } n \equiv 1 (\text{mod } 3) \\ 4e_{n-1} & \text{if } n \equiv 2 (\text{mod } 3). \end{array}$$

In the latter case, do not delete $2e_{n-1}+2e_n$. Then

$$\lambda(C) = \frac{1}{2}\left(n^2 - \frac{5}{3}n + \frac{2r}{3} + 1 - \frac{1}{3}\sqrt{6n^4 - 48n^3 + (12r+106)n^2 - (44r+48)n + (2r+3)^2}\right).$$

Proof. We must first show that C is an allowable cage; that is, the convex hull of the remaining points does not include any of the deleted points. The even points of $C_{n,4}$ have only the forms $4e_i$ and $2e_i + 2e_j$ (since the coefficient sum must be equal to 4). It is clear that we may delete any point of the form $4e_k$, since these are all extreme points of the full cage. It remains to show that points of the form $2e_{3k+1} + 2e_{3k+2}$ and $2e_{3k+1} + 2e_{3k+3}$ may be removed.

A point of the form $2e_i + 2e_j$ lies on a face of the cage; on this face, the other coordinates are all zero. Hence the only other points of C that can average to $2e_i + 2e_j$ are $4e_i$ and $4e_j$, at least one of which has already been deleted in the first step for $i = 3k+1$ and $j = 3k+2$ or $3k+3$.

We now count to determine the number of even points we have removed and the number of distinct averages we have removed.

For the full cage, $e(C_{n,4}) = \binom{n+1}{n-1} = \frac{1}{2}(n^2+n)$ and $a(C_{n,4}) = \binom{n+3}{n-1} = \frac{1}{24}(n^4 + 6n^3 + 11n^2 + 6n)$.

The precise formulas depend on the congruence class of n modulo 3. In all cases, there are initially $\binom{n}{2} = \frac{1}{2}(n^2-n)$points of the form $2e_i + 2e_j$.

For $n \equiv 0 (\text{mod } 3)$, we keep $n/3$ points of the form $4e_i$, and remove $2n/3$ such points. (Those we keep are $4e_1, 4e_4, \ldots, 4e_{n-2}$.) For each one of the points we keep, we omit exactly two points of the form $2e_i + 2e_j$, for a total of $2n/3$ of that form. Therefore we have removed a total of $\frac{4n}{3}$ points from the set of even points, and are left with

$$e = \frac{1}{2}(n^2+n) - \frac{4n}{3} = \frac{1}{2}n^2 - \frac{5}{6}n.$$

The other cases are counted similarly; the results are tabulated below.

r	e	Removed
0	$\frac{1}{2}n^2 - \frac{5}{6}n$	$\frac{4n}{3}$
1	$\frac{1}{2}n^2 - \frac{5}{6}n + \frac{1}{3}$	$\frac{4n-1}{3}$
2	$\frac{1}{2}n^2 - \frac{5}{6}n + \frac{2}{3}$	$\frac{4n-2}{3}$

That is, if r is the least nonnegative residue of n modulo 3, then $e = \frac{1}{2}n^2 - \frac{5}{6}n + \frac{r}{3}$, and we remove $\frac{4n-r}{3}$ even points from C.

We must now calculate a in each case. Removing a point of the form $4e_i$ results in the loss of exactly n averages: its average with another point of the form $4e_j$ is retained, since we have not yet removed any points of the form $2e_i + 2e_j$. Its average with a point of the form $2e_j + 2e_k$, $j, k \neq i$, is also retained, since we have both points $2e_i + 2e_j$ and

$2e_i + 2e_k$. Thus, the only averages which are lost are $(1/2)(4e_i + 4e_i)$ and averages of the form $(1/2)(4e_i + 2e_i + 2e_j)$; there are $n - 1$ of the latter. Therefore, a is reduced by exactly n when a point of the form $4e_i$ is removed.

Now assume that we have removed all points of the form $4e_i$ that we intend to remove. We must "count the cost" of removing a point of the form $2e_{3k+1} + 2e_{3k+2}$.

Consider sums of the form $(2e_{3k+1} + 2e_{3k+2}) + (2e_i + 2e_j)$. We have the following cases:

1. $i = j = 3k + 1$:
$$(2e_{3k+1} + 2e_{3k+2}) + 4e_{3k+1}$$
may only be obtained in one way; therefore, the average of these is lost.
2. $i = 3k + 1$, $j \neq 3k + 1$:
$$(2e_{3k+1} + 2e_{3k+2}) + (2e_{3k+1} + 2e_j) = 4e_{3k+1} + (2e_{3k+2} + 2e_j),$$
so the average of these is retained unless $j = 3k + 2$; thus, one more average is lost (totalling two so far).
3. $i \neq 3k + 1$, $j = 3k + 1$ has been considered in Case 2.
4. $i = 3k + 2$, $j \neq 3k + 1$ (to avoid duplicating Case 2):
$$(2e_{3k+1} + 2e_{3k+2}) + (2e_{3k+2} + 2e_j)$$
may only be obtained in one way since we no longer have a point of the form $4e_{3k+2}$; therefore, all of the averages so represented are lost. Since $j \neq 3k + 1$ and $j \neq 3k + 2$, we lose $n - 2$ more averages, bringing the total to n.
5. $i, j \notin \{3k + 1, 3k + 2\}$:
$$(2e_{3k+1} + 2e_{3k+2}) + (2e_i + 2e_j) = (2e_i + 2e_{3k+1}) + (2e_j + 2e_{3k+2}),$$
so we lose none of these averages.

We need not consider $i = j = 3k + 2$ since the point $4e_{3k+2}$ has already been deleted.

Since all averages involving $2e_{3k+1} + 2e_{3k+2}$ must be one of the above, we see that removing a point of that form results in a loss of exactly n averages.

We may similarly show that removing a point of the form $2e_{3k+1} + 2e_{3k+3}$ results in a loss of exactly n averages.

Therefore, again recalling that r denotes the least nonnegative residue of n modulo 3, we have

$$e = \frac{1}{2}n^2 - \frac{5}{6}n + \frac{r}{3} \quad \text{and} \quad a = a(C_{n,4}) - n\frac{4n - r}{3} = \frac{1}{24}(n^4 + 6n^3 + 21n^2 + 6n + 8nr).$$

Finally, we may compute $\lambda(C)$:

$$\lambda(C) = \frac{1}{2}\left(n^2 - \frac{5}{3}n + \frac{2r}{3} + 1 - \sqrt{\left(n^2 - \frac{5}{3}n + \frac{2r}{3} + 1\right)^2 - 8\left(\frac{1}{24}(n^4 + 6n^3 + 11n^2 - 26n + 8r)\right)}\right)$$
$$= \frac{1}{2}\left(n^2 - \frac{5}{3}n + \frac{2r}{3} + 1 - \frac{1}{3}\sqrt{6n^4 - 48n^3 + (12r + 106)n^2 - (44r + 48)n + (2r + 3)^2}\right). \quad \square$$

The idea of this theorem is to clip the corners, as before, but then replace certain corners in favor of deleting two other points each. In this way we take advantage of Lemma 2.7; the improvement is given in the following corollary.

Corollary 3.2. *Let C' and C be as above. Then*

$$\lim_{n\to\infty} \frac{1}{n}(\lambda(C) - \lambda(C')) = \frac{1}{\sqrt{6}} - \frac{1}{3} \approx 0.074915.$$

Example 3.3. Let $n = 4$. We use the same setup as in Example 2.1; see Fig. 5 for the cage. We have $n \equiv 1 \pmod 3$, and the points $4e_2$, $4e_3$, and $4e_4$ have been deleted, while $4e_1$ has not. The two new points that were deleted are the points $2e_1 + 2e_2$ and $2e_1 + 2e_3$, and each results in the loss of exactly four distinct averages. The value of $\lambda(C)$ is now 5.

Thus, some improvement in the estimate is certainly possible. However, as we shall see, there is a limit to how far this technique can be exploited.

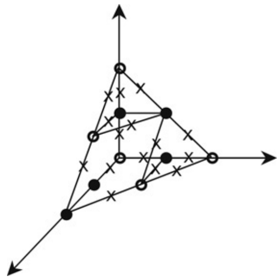

Fig. 5. Improvement (Section 3). $e(C) = 5$, $a(C) = 15$, $\lambda(C) = 5$. ∘: Deleted even point. •: Even point. x: Lost average.

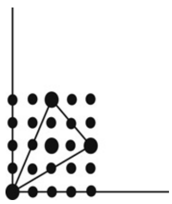

Fig. 6. Cage of the Motzkin polynomial. $e = 4$, $a = 10$, $\lambda = 4$.

4. The case $n = 3$

We now restrict ourselves to the case of three homogeneous variables.

Lemma 4.1. *Let P be a convex lattice polygon in $\mathbb{R}^2$ with i interior lattice points and b boundary lattice points (including vertices); assume that P is not a line segment. Let $2P$ be the "doubled polygon"; i.e., $2P$ is a lattice polygon similar to P with scale factor 2. Let i' and b' denote the number of lattice interior and boundary points of $2P$, respectively. Then $b' = 2b$ and $i' = 4i + b - 3$.*

Proof. Refer to Fig. 2 for an example. P has b lattice points on its boundary, and an additional b midpoints between consecutive boundary lattice points. When P is doubled, all these become boundary lattice points, and they are the only ones. Thus $b' = 2b$.

Let A be the area of P and A' the area of P'. Then from Pick's theorem, $A' = \frac{1}{2}b' + i' - 1$, and $A = \frac{1}{2}b + i - 1$. Additionally, we know that $A' = 4A$, so that $\frac{1}{2}(2b) + i' - 1 = 2b + 4i - 4$, and $i' = 4i + b - 3$. $\square$

Lemma 4.2. *Let P, b, i, b', i' be as in Lemma 4.1. Let e be the number of lattice points interior to or on the boundary of P $(e = b + i)$, and let a be the number of distinct averages of those points $(a = b' + i')$. Then $a \leq 4e - 6$.*

Proof.

$$\begin{aligned} a &= i' + b' \\ &= 4i + 3b - 3 \\ &\leq 4i + 3b - 3 + (b - 3) \quad \text{(since } P \text{ is not a line segment)} \\ &= 4(i + b) - 6 \\ &= 4e - 6. \quad \square \end{aligned}$$

Theorem 4.3. $\overline{\lambda}(C_{3,m}) = 4$ *for $m \geq 4$, m even.*

Proof. Let C be a subcage of $C_{3,m}$. If C is a line segment, then $a = 2e - 1$ and

$$\lambda(C) = \frac{2e + 1 - \sqrt{(2e+1)^2 - 8(2e-1)}}{2} = \frac{2e + 1 - \sqrt{4e^2 - 12e + 9}}{2} = 2.$$

If C is not a line segment, then we may apply Lemma 4.2 to get

$$\begin{aligned}\lambda(C) &= \frac{2e+1-\sqrt{(2e+1)^2-8a}}{2}\\ &\le \frac{2e+1-\sqrt{4e^2+4e+1-8(4e-6)}}{2}\\ &= \frac{2e+1-(2e-7)}{2}\\ &= 4.\end{aligned}$$

Since the cage corresponding to the Motzkin polynomial $(x^4y^2+x^2y^4-3x^2y^2+1)$ (see Fig. 6) has $\lambda = 4$, we conclude that $\overline{\lambda}(C_{3,m}) = 4$. □

In [1], it is shown that $P(R[x_1,\ldots,x_n]) = \infty$ by construction of a polynomial of length $t+1$ from a polynomial of length t. In particular, by following their method one can construct a polynomial with a cage in $C_{3,12}$ that has length 5. The method of modifying cages to improve the estimate on $P(R_{12}[x,y,z])$, however, will never predict the existence of this polynomial (or any polynomial of greater length). Therefore, in order to improve the lower bound estimates on Pythagoras numbers, we need to find another method, at least in the three variable case.

Unfortunately, there is some reason to suspect that the following is true, as well.

Conjecture 4.4. $\overline{\lambda}(C_{n,m}) = 2^{n-1}$ *for fixed n and all sufficiently large m.*

This brings up an interesting comparison with Pfister's famous result: $P(\mathbb{R}(x_1,\ldots,x_n)) \le 2^n$ [6]. Note that in terms of our conjecture, this would imply that $P(\mathbb{R}(x_1,\ldots,x_n)) \le 2^n \le P(\mathbb{R}_m[x_1,\ldots,x_{n+1}])$ for m sufficiently large compared to n. (Recall that $\mathbb{R}_m[x_1,\ldots,x_{n+1}]$ denotes the set of *homogeneous* polynomials of degree m in $n+1$ variables; see Definition 1.7.) This would ensure gaps between $P(\mathbb{R}(x_1,\ldots,x_n))$ and $P(\mathbb{R}_m[x_1,\ldots,x_n])$ for many values of m and n; for specific examples of such gaps, see [5] (for example).

The cage method may still be of great use in the case that n is very large compared to m, even if the conjecture holds. However, if the conjecture is true, then the cage method will hit a ceiling when the degree is sufficiently large; that is, there will be a degree beyond which the cage method cannot possibly yield the desired Pythagoras number.

References

[1] M.D. Choi, Z.D. Dai, T.Y. Lam, B. Reznick, The Pythagoras numbers of some affine algebras and local algebras, J. Reine Angew. Math. 336 (1982) 45–82.

[2] M.D. Choi, T.Y. Lam, B. Reznick, Sums of squares of real polynomials, Proc. Sym. Pure Math., Amer. Math Soc. 58.2 (1995) 103–126.

[3] D. Hilbert, Uber die Darstellung definiter Formen als Summe von Formenquadraten, Math. Ann. 32 (1888) 342–350.

[4] T.Y. Lam, Algebraic Theory of Quadratic Forms, W.A. Benjamin, 1973.

[5] D.B. Leep, C.L. Starr, Positive semidefinite polynomials in $\mathbb{R}[x, y]$ that are not a sum of squares in $\mathbb{R}(x, y)$, Proc. Amer. Math. Soc. 129 (11) (2001) 3133–3141.

[6] A. Pfister, Zur Darstellung definiter Funktionen als Summe von Quadraten, Invent. Math. 4 (1967) 229–237.

[7] B. Reznick, Extremal PSD Forms with few terms, Duke Math. J. 45 (1978) 363–374.

[8] B. Reznick, Forms derived from the arithmetic-geometric inequality, Math. Ann. 283 (1989) 431–464.

[9] G.M. Ziegler, Lectures on Polytopes (Lecture 2), Springer-Verlag, 1995.